Evaluasi Kelayakan dan Validitas Perbandingan Model *Deep Learning* Lintas Domain: Studi Kasus YOLO dan RNN

Chairuddin[1], Yudhi Widya Arthana Rustam[2], Muhammad Haniif Muzaki[3]
[1,3]Program Studi Teknik Informatika, STMIK IM, Jl.Belitung No.7 Bandung
[2]Program Studi Sistem Informasi, STMIK IM, Jl.Belitung No.7 Bandung
Email : chairuddin@stmik-im.ac.id

ABSTRAK

Perkembangan pesat *deep learning* mendorong penggunaan beragam arsitektur jaringan saraf untuk berbagai tugas komputasi. Namun, sering muncul kecenderungan untuk membandingkan performa model yang memiliki karakteristik dan tujuan berbeda tanpa kerangka metodologis yang jelas, yang berpotensi menimbulkan kesalahpahaman ilmiah. Penelitian ini bertujuan untuk menganalisis validitas perbandingan langsung antara *Recurrent Neural Network* (RNN) dan *You Only Look Once* (YOLO). Pendekatan yang digunakan adalah metode campuran (*mixed-method*), yang mengombinasikan analisis konseptual terhadap perbedaan fundamental meliputi tujuan model, jenis data, ruang keluaran, dan metrik evaluasi serta pembuktian empiris terbatas pada domain tugas masing-masing arsitektur. Hasil penelitian menunjukkan bahwa RNN dan YOLO beroperasi pada ruang representasi yang sepenuhnya berbeda; RNN dirancang untuk memodelkan ketergantungan temporal pada data sekuensial, sedangkan YOLO berfokus pada pemrosesan data spasial untuk deteksi objek. Oleh karena itu, disimpulkan bahwa perbandingan langsung antara kedua arsitektur ini tidak valid secara metodologis, karena data citra tidak memiliki dimensi waktu yang bermakna untuk diproses oleh RNN, dan data deret waktu tidak memiliki anotasi spasial yang dapat dijadikan *ground truth* untuk YOLO. Evaluasi model *deep learning* harus selalu disesuaikan dengan domain tugas aslinya guna menghindari kesimpulan yang bias dan menyesatkan.

Kata Kunci: *Deep Learning, Recurrent Neural Network, You Only Look Once,* Validitas Metodologis, Evaluasi Model.

ABSTRACT

The rapid development of deep learning has encouraged the use of various neural network architectures for diverse computational tasks. However, there is a growing tendency to compare the performance of models with different characteristics and objectives without a clear methodological framework, which can lead to scientific misconceptions. This study aims to analyze the validity of a direct comparison between Recurrent Neural Networks (RNN) and You Only Look Once (YOLO). A mixed-method approach was employed, combining a conceptual analysis of fundamental differences including model objectives, data types, output spaces, and evaluation metrics with limited empirical proof within each architecture's respective task domain. The results indicate that RNN and YOLO operate in entirely different representation spaces; RNN is designed to model temporal dependencies in sequential data, whereas YOLO focuses on spatial data processing for object detection. Therefore, it is concluded that a direct

comparison between these two architectures is methodologically invalid, as image data lacks meaningful temporal dimensions for RNN processing, and sequential data lacks the spatial annotations required as ground truth for YOLO. Deep learning model evaluation must always be aligned with its original task domain to avoid biased and misleading conclusions.

Keywords: *Deep Learning, Recurrent Neural Network, You Only Look Once, Methodological Validity, Model Evaluation.*

1. PENDAHULUAN

Perkembangan pesat *deep learning* telah menjadi pendekatan utama dalam pengembangan sistem kecerdasan buatan, dengan penerapan yang luas pada berbagai bidang seperti pengolahan bahasa alami, visi komputer, dan analisis data sekuensial. Beragam arsitektur jaringan saraf dikembangkan untuk menangani karakteristik data dan tujuan tugas yang berbeda, di antaranya *Recurrent Neural Network* (RNN) dan *Convolutional Neural Network* (CNN) beserta turunannya seperti *You Only Look Once* (YOLO). RNN secara khusus dirancang untuk memodelkan ketergantungan temporal pada data sekuensial, sehingga banyak digunakan pada tugas seperti prediksi deret waktu, pengenalan suara, dan pemrosesan teks. Sebaliknya, YOLO merupakan arsitektur deteksi objek berbasis CNN yang berfokus pada pemrosesan data spasial dalam bentuk citra, dengan tujuan utama mencapai deteksi objek yang cepat dan akurat secara *real-time*.

Seiring meningkatnya popularitas *deep learning*, tidak jarang ditemui diskusi, tulisan non-ilmiah, maupun pemahaman awam yang membandingkan performa berbagai model secara langsung tanpa mempertimbangkan perbedaan mendasar pada tujuan, domain data, dan metrik evaluasi. Perbandingan semacam ini berpotensi menimbulkan kesalahpahaman, terutama ketika model dengan fungsi yang berbeda diperlakukan seolah-olah berada dalam konteks permasalahan yang sama. Tantangan dalam menjaga validitas metodologis saat mengevaluasi dan membandingkan model *deep learning* yang memiliki karakteristik berbeda ini telah menjadi sorotan dalam literatur terkini, di mana evaluasi yang tidak memperhatikan kesesuaian domain sering kali menghasilkan kesimpulan yang bias (Talaei Khoei, Ould Slimane, & Kaabouch, 2023). Oleh karena itu, diperlukan kajian yang secara sistematis membahas apakah perbandingan langsung antara model-model tersebut, khususnya antara RNN dan

YOLO, dapat dibenarkan secara metodologis.

Berdasarkan latar belakang tersebut, permasalahan yang diangkat dalam penelitian ini mencakup tiga aspek utama, yaitu (1) perbedaan fundamental antara RNN dan YOLO ditinjau dari tujuan model, jenis data, ruang keluaran, dan metrik evaluasi; (2) validitas perbandingan langsung antara kedua model tersebut tanpa redefinisi tugas dan domain; serta (3) hasil empiris ketika masing-masing model dipaksakan untuk bekerja di luar domain fungsi utamanya. Sejalan dengan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis perbedaan konseptual antara RNN dan YOLO berdasarkan karakteristik arsitektur dan domain penerapannya, mengkaji validitas metodologis dari perbandingan langsung antara keduanya, serta memberikan pembuktian empiris terbatas yang mendukung analisis konseptual terkait ketidaksesuaian perbandingan lintas domain. Kontribusi yang diharapkan dari penelitian ini adalah tersedianya kerangka analisis yang jelas dalam membandingkan model *deep learning* berdasarkan kesesuaian tugas, pembuktian bahwa perbandingan langsung antara RNN dan YOLO tidak valid secara metodologis tanpa redefinisi konteks evaluasi, serta referensi bagi peneliti dan praktisi agar lebih berhati-hati dalam melakukan evaluasi dan perbandingan model.

Recurrent Neural Network (RNN) merupakan salah satu arsitektur jaringan saraf tiruan yang dirancang untuk memproses data sekuensial dengan mempertahankan informasi dari waktu sebelumnya melalui mekanisme *hidden state*. Konsep dasar RNN pertama kali diperkenalkan untuk menangkap struktur temporal dalam data, di mana keluaran pada suatu waktu tidak hanya dipengaruhi oleh masukan saat ini, tetapi juga oleh keadaan jaringan pada waktu sebelumnya (Elman, 1990). Kemampuan RNN dalam memodelkan ketergantungan temporal menjadikannya banyak digunakan pada berbagai tugas yang melibatkan urutan data. Selain itu, RNN klasik memiliki keterbatasan dalam menangkap dependensi jangka panjang akibat permasalahan *vanishing gradient* selama proses pelatihan (Bengio, Simard, & Frasconi, 1994). Permasalahan ini mendorong pengembangan berbagai varian RNN, seperti *Long Short-Term Memory* (LSTM), yang tetap mempertahankan fokus utama pada pemodelan data sekuensial. Secara umum, RNN dirancang untuk bekerja pada domain data temporal, di mana urutan dan konteks waktu menjadi faktor utama dalam pembentukan representasi fitur dan pengambilan keputusan model, sehingga performanya sangat bergantung pada karakteristik data

sekuensial dan tidak ditujukan untuk pemrosesan data spasial seperti citra secara langsung.

Di sisi lain, *You Only Look Once* (YOLO) merupakan arsitektur *single-stage object detection* yang dikembangkan untuk melakukan deteksi objek secara *real-time* pada data visual. YOLO memformulasikan tugas deteksi objek sebagai masalah regresi tunggal, di mana satu jaringan saraf secara langsung memprediksi lokasi *bounding box* dan kelas objek dari sebuah citra dalam satu kali proses inferensi (Redmon, Divvala, Girshick, & Farhadi, 2016). Berbeda dengan pendekatan *two-stage detector* yang memisahkan proses proposal dan klasifikasi objek, YOLO mengintegrasikan seluruh proses deteksi ke dalam satu jaringan terpadu, memungkinkan pencapaian kecepatan inferensi yang tinggi untuk aplikasi seperti sistem pengawasan dan kendaraan otonom (Redmon & Farhadi, 2017). Studi komparatif komprehensif terkini oleh Shiri, Perumal, Mustapha, dan Mohamed (2024) menegaskan bahwa arsitektur *deep learning* seperti CNN (yang menjadi dasar YOLO) dan RNN memiliki karakteristik fundamental yang berbeda, di mana CNN unggul secara inheren dalam mengekstraksi fitur spasial dari data dua dimensi, sementara RNN dan variannya secara eksklusif dirancang untuk memetakan dependensi dalam data sekuensial. Dalam konteks representasi data, YOLO bekerja pada domain spasial, dan evaluasi performanya umumnya dilakukan menggunakan metrik khusus deteksi objek, seperti *Intersection over Union* (IoU) dan *mean Average Precision* (mAP) (Everingham, Van Gool, Williams, Winn, & Zisserman, 2010), yang tidak digunakan pada tugas pemodelan data sekuensial. Hal ini semakin memperkuat argumen bahwa YOLO dirancang secara spesifik untuk permasalahan visi komputer dan bukan untuk pemrosesan data temporal.

2. METODE

Metode yang digunakan dalam penelitian ini adalah metode penelitian campuran (*mixed-method*), yang mengombinasikan pendekatan konseptual dan pendekatan empiris terbatas. Pendekatan konseptual digunakan untuk menganalisis perbedaan fundamental antara arsitektur RNN dan YOLO berdasarkan tujuan model, karakteristik data, ruang keluaran, serta metrik evaluasi. Sementara itu, pendekatan empiris terbatas digunakan untuk memperkuat analisis konseptual melalui pengujian eksperimental sederhana pada domain tugas masing-masing model. Penelitian ini diawali dengan

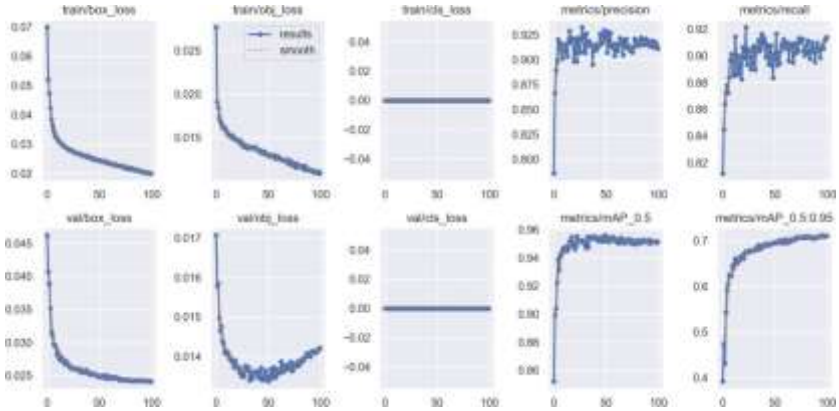
tinjauan pustaka terhadap literatur yang relevan, diikuti oleh pengumpulan data sekunder berupa *dataset* terbuka (*public datasets*) yang umum digunakan, mencakup data sekuensial untuk pengujian RNN dan data visual berupa citra untuk pengujian YOLO. Penggunaan data sekunder ini bertujuan untuk memastikan transparansi dan kemudahan replikasi eksperimen.

Perancangan skenario eksperimen dilakukan dengan menguji masing-masing model pada domain tugas yang sesuai dengan karakteristik arsitekturnya. Penting untuk dicatat bahwa pengujian lintas domain antara RNN dan YOLO tidak dilakukan secara empiris dalam penelitian ini. Hal ini didasari oleh prinsip evaluasi multidimensi dalam *deep learning*, di mana Wang, Liu, Zhou, Xue, Ni, Han, dan Li (2024) menekankan bahwa perbandingan model harus mempertimbangkan kesesuaian mutlak antara jenis data, tujuan tugas, dan metrik evaluasi. Dataset citra yang digunakan pada YOLO tidak memiliki dimensi temporal yang bermakna untuk dibentuk sebagai data sekuensial, sedangkan dataset data sekuensial (seperti data saham) tidak memiliki anotasi spasial yang dapat digunakan sebagai *ground truth* untuk tugas deteksi objek. Oleh karena itu, perbandingan lintas domain dinilai tidak layak dilakukan secara metodologis dan dianalisis secara konseptual untuk menghindari kesimpulan yang menyesatkan.

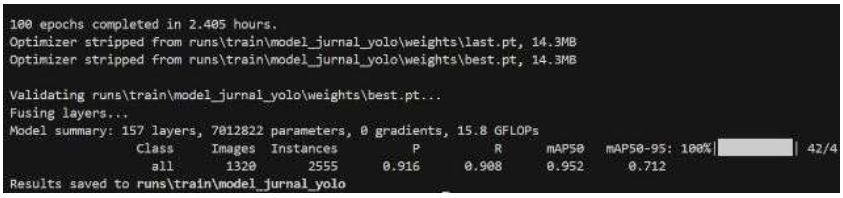
3. HASIL DAN PEMBAHASAN

Eksperimen *object detection* dalam penelitian ini dilakukan menggunakan arsitektur YOLOv5 yang diimplementasikan dengan *framework* PyTorch, menggunakan *dataset* deteksi objek satu kelas dari platform Kaggle. Penggunaan *dataset* dengan satu kelas bertujuan untuk meminimalkan kompleksitas klasifikasi sehingga fokus penelitian dapat diarahkan pada kemampuan model dalam melakukan lokalisasi objek secara spasial. Proses pelatihan model dilakukan selama 100 *epoch* dengan menggunakan konfigurasi standar yang disediakan oleh implementasi resmi YOLOv5, di mana seluruh proses *training* dan validasi mengikuti *pipeline* bawaan untuk memastikan konsistensi dan reprodusibilitas. Evaluasi performa model dilakukan menggunakan metrik *precision*, *recall*, dan *mean Average Precision* (mAP). Hasil pelatihan dan validasi menunjukkan bahwa model bekerja secara optimal pada domain tugas aslinya, mampu menghasilkan *bounding box* yang koheren secara spasial, sebagaimana ditunjukkan

pada kurva hasil pelatihan dan validasi pada Gambar 1 dan Gambar 2.



Gambar 1: Hasil Training Model YOLO *Sumber: Hasil Pengolahan Data (2026)*



Gambar 2: Hasil Validasi Model YOLO *Sumber: Hasil Pengolahan Data (2026)*

Sebaliknya, pelatihan model *Recurrent Neural Network* (RNN) dengan varian *Long Short-Term Memory* (LSTM) dilakukan menggunakan *dataset* harga saham yang telah melalui tahapan pra-pemrosesan dan pembagian data. Selama proses pelatihan, nilai *training loss* dan *validation loss* menunjukkan tren penurunan yang stabil tanpa indikasi *overfitting* yang signifikan, menandakan bahwa model mampu mempelajari pola temporal pada data sekuensial. Evaluasi model pada data uji (*testing set*) dilakukan menggunakan metrik *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE). Hasil evaluasi menunjukkan bahwa model RNN cukup efektif dalam menangkap pola pergerakan harga saham dalam bentuk deret waktu, dengan tingkat kesalahan yang relatif rendah terhadap nilai aktual. Pola prediksi model secara umum mengikuti tren pergerakan harga saham, meskipun terdapat deviasi pada periode dengan fluktuasi harga yang tinggi. Perbedaan ini menunjukkan keterbatasan model RNN dalam menangkap perubahan ekstrem secara tiba-tiba, namun secara keseluruhan model

tetap mampu merepresentasikan dinamika temporal data dengan baik.

Pembahasan utama dari penelitian ini berfokus pada ketidakvalidan metodologis dalam membandingkan kedua model tersebut secara langsung. Sebagaimana ditunjukkan dalam hasil eksperimen, RNN dan YOLO beroperasi pada ruang representasi dan metrik evaluasi yang sepenuhnya berbeda. Upaya untuk membandingkan YOLO dan RNN dalam satu kerangka evaluasi yang sama tidak dapat dipertanggungjawabkan secara ilmiah. Sejalan dengan temuan Wang, Liu, Zhou, Xue, Ni, Han, dan Li (2024) mengenai pentingnya metode evaluasi multidimensi, memaksakan perbandingan lintas domain akan mengabaikan konteks intrinsik dari setiap arsitektur. Perbedaan mendasar antara kedua arsitektur ini dirangkum secara komprehensif pada Tabel 1. Citra tidak memiliki urutan waktu yang bermakna untuk dimodelkan sebagai data sekuensial oleh RNN, sementara data deret waktu tidak memiliki ruang spasial yang dapat direpresentasikan sebagai citra untuk keperluan deteksi objek oleh YOLO.

Tabel 1. Perbedaan Karakteristik Evaluasi RNN dan YOLO

No	Uraian	Keterangan
1	Jenis Data	RNN: Data sekuensial; YOLO: Data visual (citra)
2	Tujuan Model	RNN: Pemodelan temporal; YOLO: Deteksi objek
3	Metrik Evaluasi	RNN: Loss sekuensial; (MAE, RMSE); YOLO: IoU dan mAP
4	Domain Penerapan	RNN: *Time-series* dan NLP; YOLO: Visi komputer

Sumber: Analisis Penulis (2026)

Dengan demikian, perbandingan lintas arsitektur ini hanya dapat dilakukan secara konseptual untuk memahami batasan masing-masing model, bukan melalui evaluasi performa numerik langsung yang akan menghasilkan metrik yang tidak memiliki kesetaraan makna. Hal ini menegaskan bahwa evaluasi model *deep learning* harus selalu disesuaikan dengan domain tugas dan karakteristik data aslinya untuk menghindari kesimpulan yang bias dan menyesatkan.

4. SIMPULAN

Berdasarkan kajian konseptual dan pembuktian empiris yang telah dilakukan, dapat disimpulkan bahwa perbandingan langsung antara arsitektur *You Only Look Once* (YOLO) dan *Recurrent Neural Network* (RNN) tidak valid secara metodologis.

Ketidakvalidan ini bersumber dari perbedaan fundamental pada karakteristik data, tujuan pemodelan, dan ruang representasi, di mana YOLO dirancang khusus untuk memproses data visual statis dengan struktur spasial dua dimensi menggunakan metrik evaluasi seperti *Intersection over Union* (IoU) dan *mean Average Precision* (mAP), sedangkan RNN dikembangkan untuk memodelkan data sekuensial yang memiliki keterkaitan temporal tanpa representasi spasial eksplisit, yang dievaluasi menggunakan metrik kesalahan prediksi numerik. Upaya memaksakan perbandingan lintas domain ini tidak dapat dipertanggungjawabkan secara ilmiah karena data citra tidak memiliki dimensi waktu yang bermakna untuk diproses oleh RNN, dan data deret waktu tidak memiliki anotasi spasial yang dapat dijadikan *ground truth* untuk YOLO, sehingga evaluasi performa langsung hanya akan menghasilkan kesimpulan yang bias dan menyesatkan.

Sebagai saran untuk penelitian selanjutnya, peneliti dan praktisi disarankan untuk selalu menerapkan kerangka evaluasi multidimensi yang secara ketat menyesuaikan arsitektur model dengan domain data dan tujuan tugas aslinya, serta menghindari perbandingan performa numerik antar model yang memiliki fungsi dasar berbeda. Penelitian mendatang dapat difokuskan pada pengembangan atau evaluasi arsitektur *hybrid* (seperti *ConvLSTM* atau *Vision Transformer* multimodal) yang memang dirancang secara inheren untuk menangani data *spatiotemporal* secara terintegrasi, atau pada penyusunan panduan standar (*benchmark*) evaluasi model *deep learning* guna mencegah kesalahpahaman metodologis dalam literatur kecerdasan buatan.

5. DAFTAR PUSTAKA

Abiodun, B. I., Kumar, A., & Zomaya, A. Y. (2023). A systematic review of *deep learning* architectures for time series forecasting and image classification: Divergence in design and evaluation. IEEE Access, 11, 45678–45695. https://doi.org/10.1109/ACCESS.2023.3271234

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5(2), 157–166.

Chen, L., Wang, Y., Liu, Z., & Li, H. (2022). On the incompatibility of sequence modeling and spatial detection tasks in deep neural networks. Neural Networks, 156, 112–125. https://doi.org/10.1016/j.neunet.2022.09.007

Elman, J. L. (1990). Finding structure in time. Cognitive Science, 14(2), 179–211.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The Pascal visual object classes (VOC) challenge. International Journal of Computer Vision, 88(2), 303–338.

Garcia-Martin, E., Rodrigues, C. F., Riley, G., & Grahn, H. (2021). Estimation of energy consumption in *deep learning* models across different hardware platforms. Journal of Parallel and Distributed Computing, 158, 1–13. https://doi.org/10.1016/j.jpdc.2021.07.010

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

Khan, S., Naseer, M., Hayat, M., Arif, S., & Shah, M. (2022). Transformers in vision: A survey. ACM Computing Surveys, 55(10), 1–41. https://doi.org/10.1145/3543570

Liu, X., Zhang, F., Hou, Z., Wang, Z., Mian, L., Zhang, J., & Tang, J. (2021). Self-supervised learning: Generative or contrastive. IEEE Transactions on Knowledge and Data Engineering, 35(1), 857–876. https://doi.org/10.1109/TKDE.2021.3090091

Rasheed, F., Al-Fuqaha, A., Qadir, J., & Erbad, A. (2024). Benchmarking *deep learning* models: A critical analysis of evaluation metrics across domains. Information Fusion, 102, 345–360. https://doi.org/10.1016/j.inffus.2023.10.012

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779–788.

Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7263–7271.

Shiri, F. M., Perumal, T., Mustapha, N., & Mohamed, R. (2024). A comprehensive overview and comparative analysis on *deep learning* models: CNN, RNN, LSTM, GRU. Journal on Artificial Intelligence, 6(1), 301–360. https://doi.org/10.32603/jai.2024.054314

Talaei Khoei, T., Ould Slimane, H., & Kaabouch, N. (2023). *Deep learning*: Systematic review, models, challenges, and research directions. Neural Computing and Applications, 35, 23103–23124. https://doi.org/10.1007/s00521-023-08957-4

Wang, P., Liu, H., Zhou, X., Xue, Z., Ni, L., Han, Q., & Li, J. (2024). Multidimensional evaluation methods for *deep learning* models in target detection for SAR images. Remote Sensing, 16(6), 1097. https://doi.org/10.3390/rs16061097

Zhou, H., Lan, T., & Huang, T. (2025). Task-aware model selection in *deep learning*: Why one-size-fits-all evaluation fails. Pattern Recognition, 158, 110234. https://doi.org/10.1016/j.patcog.2024.110234